AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

0096-3445/19/$12.00

http://dx.doi.org/10.1037/xge0000552

Interattribute Evaluation Theory

Ioannis Evangelidis
Bocconi University

Stijn M. J. van Osselaer
Cornell University

In this article we advance a theory that describes how people evaluate attribute values. We propose that evaluations involve a target and a reference value. Evaluators first seek a reference value on the target attribute (e.g., an average value or another stimulus's value on that same attribute). However, in the absence of same-attribute information, evaluators may instead rely on the target stimulus's own value on another attribute and make an evaluation about the target in one of two ways. First, the individual may compare the target attribute value to the stimulus's value on a reference attribute. The evaluator is more likely to engage in an interattribute comparison when the target attribute value is relatively evaluable and compatible with the reference value. Second, the individual may infer the magnitude of the target value based on his or her judgment about the extremity (e.g., the goodness or badness) of the stimulus's value on a reference attribute and the perceived correlation between the target attribute and the reference attribute. The evaluator is more likely to make an inference about the target value based on the reference when the target is low in evaluability and is less compatible with the reference value. Two attribute values are considered to be more compatible when their scale format is more similar. We provide support for our framework in 14 studies.

Keywords: interattribute evaluation theory, attribute evaluations, attribute value judgments, magnitude judgments, evaluability theory

How large is the amount of $100? How good is an SAT score of 1200? How large is the number of 800 people left homeless by a natural disaster? Evaluating attribute values is a cornerstone of virtually every judgment or decision that humans make. In this paper we advance and empirically test a theory of how people evaluate attribute values.

We propose that evaluations typically involve a "target" and a "reference" value. The target is the value under evaluation. The reference is an additional value (or point) that the individual considers when making an evaluative judgment about the target. The reference may be another value on the same attribute, but also another value on another attribute altogether. Prior research provides robust evidence that individuals commonly compare the target stimulus's target attribute value with another (reference) value on the same attribute—as in joint evaluation (Hsee, 1996; Hsee, Loewenstein, Blount, & Bazerman, 1999; Hsee & Zhang, 2010). For example, when a layperson is exposed to information about the number of assists of a basketball player, she can use that information as reference to evaluate the number of assists handed out by another player.

Whereas prior research provides ample evidence for evaluations as a function of within-attribute comparisons, it remains silent on how individuals evaluate attribute values when another value on the same attribute is not available. In this article, we posit that, under such circumstances, an evaluator may often rely on one of two additional processes that both involve the use of a reference value associated with another attribute. In our theoretical analysis we assume that the additional (reference) attribute is relatively evaluable. (Normatively, we should not expect the evaluator to use information about a reference attribute when she cannot make sense of that information.)

The first process is an interattribute comparison process where the individual compares the target attribute value to the stimulus's value on another relevant (reference) attribute of that same stimulus. Her evaluation is based on an assessment of the target value in the context of the reference value (i.e., "given reference value X, how high/good is Y?"). We conjecture that the evaluator's tendency to engage in interattribute comparisons is contingent on two factors: the evaluability of the target attribute and the compatibility of the target value with the reference value. First, and perhaps intuitively, the evaluator should be more likely to compare the target value to the stimulus's value on another (reference) attribute when the target is relatively evaluable. Second, the evaluator should be more likely to compare the target value to the reference value when attribute value compatibility is relatively high. We

Ioannis Evangelidis, Department of Marketing, Bocconi University; Stijn M. J. van Osselaer, Samuel Curtis Johnson Graduate School of Management, Cornell SC Johnson College of Business, Cornell University.

We truthfully report all measures and conditions. We did not exclude any data from our analyses. We acknowledge that our predictions and theoretical framework were revised through the review process. Part of this research was presented at the Association for Consumer Research Conference in Dallas (October 2018) and at the Annual Meeting of the Society for Judgment and Decision-Making in New Orleans (November 2018). Data can be accessed at https://osf.io/gc47y/.

Correspondence concerning this article should be addressed to Ioannis Evangelidis, Department of Marketing, Bocconi University, Via Roentgen 1, Milan 20136, Italy. E-mail: ioannis.evangelidis@unibocconi.it

define attribute value compatibility as the degree of similarity of the scales of the two values—target and reference. For instance, attribute value compatibility is higher when the two attributes are measured on the same or on a similar scale (e.g., when a given value on one attribute is a possible or valid value on the other attribute) or when their values have a similar format (e.g., when both values are integers). Attribute-value compatibility should impact the ease of making an (interattribute) comparison (i.e., comparison fluency; Simonson, Bettman, Kramer, & Payne, 2013). For example, imagine the case of a TV spectator who just tuned in to an NBA game broadcast. She is presented with information about the number of points and assists by a single player ("Player A"). We propose that, when evaluating the number of assists handed out by Player A, this spectator will compare that number to the number of points scored by the same player since the two values are relatively compatible. (Their values are compatible because the number of assists is within the possible range of number of points and vice versa, while both attributes assume only integer values). Therefore, in this example, our theory implies that the same target attribute value (e.g., number of assists) will be evaluated differently depending on the stimulus's score on the reference attribute (e.g., number of points). Holding the actual number of assists fixed, the number of assists will be seen as larger when the number of points is low versus high.

The second process is an interattribute inference process. The individual may make an inductive inference about the known value of the target attribute using a process akin to the inferential imputation process used in the estimation of missing attribute information (Jaccard & Wood, 1988; Kardes, Posavac, & Cronley, 2004). Specifically, the evaluator may infer the magnitude of the target value based on her judgment about the extremity of the stimulus's value on a reference attribute (i.e., "how high/good is reference value X?") and the perceived correlation between the target attribute and the reference attribute (i.e., "how strongly does the reference attribute correlate with the target attribute?"). The evaluator should be more likely to make an inference about the target depending on the evaluability of the target attribute and the attribute value compatibility of the two attribute values. First, the evaluator should be more likely to make an inference about the target based on the reference value when the target attribute is relatively hard to evaluate, thus being less evaluable than the reference attribute. Second, the evaluator should be more likely to make an inference when the two values are relatively incompatible. Attribute values are incompatible when measured on a different scale because a given value on one attribute is often not a possible (or valid) value on the other attribute, or when their values have a different format (e.g., when one attribute assumes decimal values, while the other attribute assumes integer values). For example, imagine the case of another TV spectator who just tuned in to a different NBA game broadcast. She is presented with information about the number of points and the efficiency score of a player ("Player B"). For most, efficiency should be relatively hard to evaluate (and harder to evaluate than the number of assists in our previous example). Furthermore, the spectator cannot always directly compare the efficiency score to the number of points since the two attribute values are often incompatible (because, in contrast to the number of points, efficiency can assume negative and/or decimal values). Thus, in this example, we argue that the evaluation of the efficiency score will most likely be driven by the magnitude of the number of points and the perceived correlation between the two attributes. That is, holding the actual efficiency score fixed, the efficiency score will be perceived to be larger as the number of points increases and the two attributes are seen as positively correlated.

To summarize, we propose that the evaluation of a target value is influenced by a reference value which may be either another value on the same attribute or another value on another attribute altogether. We presume that the evaluator will tend to rely most on same-attribute information when available because same-attribute information is more diagnostic than information from other attributes. However, in the absence of information about another value on the target attribute, the individual may instead rely on the stimulus's value on another (reference) attribute and make an evaluation about the target in one of two ways. First, the individual may compare the target attribute value to the stimulus's value on a reference attribute. The evaluator should be more likely to engage in an interattribute comparison when the target attribute is relatively evaluable and the two values are relatively compatible. Second, the individual may infer the magnitude of the target value based on her judgment about the extremity (e.g., the goodness or badness) of the stimulus's value on a reference attribute and the perceived correlation between the target attribute and the reference attribute. The evaluator is more likely to make an inference about the target based on the reference when the target is hard to evaluate and the two values are relatively incompatible. Our predictions about the two interattribute evaluation processes are illustrated in Figure 1. The two processes produce directionally opposite effects on attribute evaluations. When the evaluator compares the target to the reference, identical targets may be perceived to be smaller (or worse) as reference values increase. In contrast, when the evaluator makes an inference about the target based on the reference, identical targets may be perceived to be larger (or better) as reference values increase.

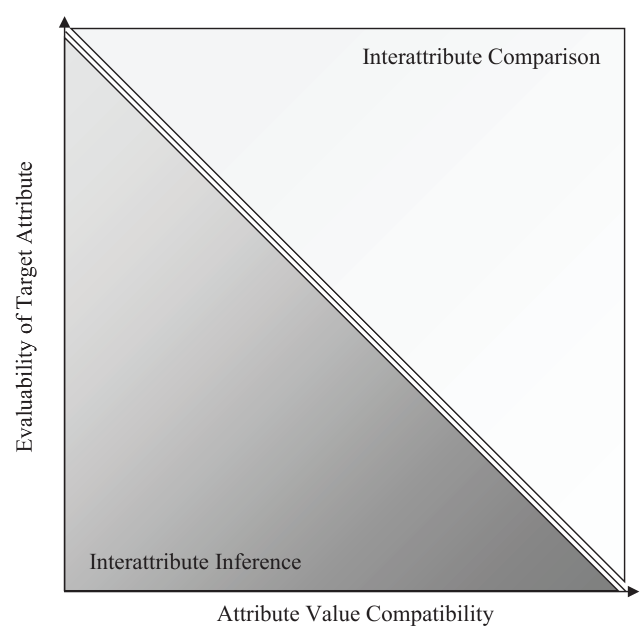

Figure 1. Attribute value compatibility, evaluability of the target, and evaluation process.

Overview of Studies

We test our theory in 14 studies (see Table 1 for an overview). Given that prior research provides ample evidence for within-attribute comparisons (for a review see Hsee et al., 1999), our investigation is focused on providing evidence for the two additional processes that employ the use of values across different attributes. Importantly, we also provide process evidence for our overall framework by demonstrating that the impact of each of these processes on evaluations is contingent on the availability of information about other values on the target attribute. We further show that the strength of each of the two focal processes is contingent on the evaluability of the target and the degree of compatibility of the two attribute values—target and reference.

Studies 1a–1c provide evidence for the interattribute comparison process across different contexts. We show that, holding target values fixed, targets are perceived to be smaller (or worse) as reference values increase. Consistent with our framework, in Study 2 we show that this effect is moderated when same-attribute information is made available. In specific, we provide some participants with a reference point within the target attribute (i.e., the average value that the target assumes), while others are asked to evaluate the target in joint (as opposed to separate) evaluation mode. We find that the impact of the interattribute comparison on evaluations of the target value diminishes when we provide participants with same-attribute information, presumably because the same-attribute information is perceived to be more diagnostic than information from another attribute.

Next, we turn our attention to the interattribute inference process. Studies 3a–3c provide evidence for this process across different contexts. We show that, holding target values fixed, targets are perceived to be larger (or better) as reference values increase. Studies 4 and 5 provide further process evidence for our theoretical framework. In Study 4, we show that judgments about the attractiveness of the reference value correlate positively with judgments about the attractiveness of the target value consistent with the interattribute inference process. Furthermore, Study 4 also shows that the impact of the interattribute inference process on evaluations is attenuated in joint compared to separate evaluation mode, presumably because—as previously argued—same-attribute information is seen as more diagnostic relative to information from another attribute. In Study 5, we test our hypothesis that the extent to which the two attributes are seen as correlated moderates the strength of the interattribute inference process. The impact of the reference value on target evaluations should be stronger when the two attributes are perceived to be highly correlated. For instance, the impact of the reference value on the target should be stronger when both attributes are about academic performance, rather than when one attribute is about academic performance, while the other is about athletic performance. The data of Study 5 provide support for this prediction.

In Studies 6–9 we provide further evidence for our theoretical framework by showing that the two evaluation processes are contingent on the target's evaluability and on the degree of compatibility of the two values. We show that the prevalence of the comparison process increases—while the prevalence of the inference process decreases—when the evaluability of the target and attribute value compatibility increase.

Finally, in Study 10 we extend our findings to preference. Similar to previous studies, we manipulate whether participants evaluate a target value based on the interattribute comparison or inference process. We further show that attribute-level evaluations influence participants' overall preference for the target.

Studies 1a–1c

Participants in Studies 1a–1c were randomly assigned to one of two conditions. Participants in both conditions were asked to evaluate the same (target) attribute value of a given stimulus, while we manipulated the level of a second (reference) attribute of the same stimulus. Consistent with an interattribute comparison process, we find robust evidence that identical target values are perceived to be smaller when reference values are high as opposed to low (see Table 1).

Study 1a

Participants in Study 1a (N = 200: 52% female; M_{age} = 37.3 years, MTurk) were asked to imagine that they were watching an NBA game. Participants were presented with a table featuring information about a player's stats (labeled "Player A"). In the low (high) points condition, the player had scored 5 (35) points and he had handed out 4 assists. We asked participants, "In your opinion, how high is the number of assists handed out by Player A?" (slider bar: 0 = *very low* to 100 = *very high*). The same number of assists was perceived to be higher when the player had scored a low (M = 46.18, SD = 22.02) compared to high number of points (M = 33.97, SD = 18.78), $t(198) = 4.22$, $p < .001$.

Study 1b

In a follow-up study, we wanted to examine whether this effect was limited to the particular target attribute (i.e., assists) in Study 1a or generalizes to other attributes. Participants in Study 1b (N = 403: 54% male; M_{age} = 34.5 years, MTurk) were asked to imagine that they were watching an NBA game. Participants were presented with a table featuring information about a player's stats (labeled "Player A"). In the low (high) points conditions the player had scored 6 (36) points. In the assists (rebounds) conditions, the player had 5 assists (rebounds). Participants were asked, "In your opinion, how high is the number of assists handed out (rebounds grabbed) by Player A?" (slider bar: 0= *very low* to 100 = *very high*). We found a significant effect of number of points, $F(1, 399) = 21.01$, $p < .001$. The same number of assists (or rebounds, depending on the condition) was perceived to be higher when the player had scored a low (M = 50.61, SD = 22.63) compared to high number of points (M = 40.39, SD = 22.09). This effect was not moderated by the target attribute (assists vs. rebounds), $F(1, 399) = 0.30$, $p > .58$.

Study 1c

Participants in Study 1c (N = 201: 60% female; M_{age} = 35.8 years, MTurk) were asked to imagine that an earthquake had taken place in Asia. Participants were presented with a table featuring information about the earthquake's impact. In the low (high) fatalities condition the number of people killed by the earthquake was 40 (3,000). The number of people left homeless by the

Table 1
Summary of Results

Reference	Target	Moderator	Mean (*SD*)
Study 1a—NBA ($N = 200$)			
Points: 5	Assists: 4		46.18 (22.02)
Points: 35	Assists: 4		33.97 (18.78)
Main Findings: The same number of assists was perceived to be higher when the player had scored a low compared to high number of points, $t(198) = 4.22$, $p < .001$.			
Study 1b—NBA ($N = 403$)			
Points: 6	Assists: 5		51.13 (21.58)
Points: 36	Assists: 5		41.09 (22.79)
Points: 6	Rebounds: 5		50.11 (23.71)
Points: 36	Rebounds: 5		39.68 (21.46)
Main Findings: The same number of assists or rebounds was perceived to be higher when the player had scored a low compared to high number of points, $F(1, 399) = 21.01$, $p < .001$.			
Study 1c—Earthquake ($N = 201$)			
Killed: 40	Homeless: 50		4.64 (2.21)
Killed: 3000	Homeless: 50		3.39 (2.20)
Main Findings: The same number of homeless people was perceived to be larger when the number of fatalities was low compared to high, $t(199) = 4.03$, $p < .001$.			
Study 2–NBA ($N = 402$)			
Points: 5	Assists: 4		43.71 (20.63)
Points: 33	Assists: 4		31.06 (22.46)
Points: 5	Assists: 4	Reference Point	53.43 (24.51)
Points: 33	Assists: 4		55.54 (27.61)
Points: 5	Assists: 4	Joint Evaluation	39.51 (23.07)
Points: 33	Assists: 4		39.52 (23.07)
Main Findings: In separate evaluation, when no same-attribute reference point was provided, the same number of assists was perceived to be higher when the player had scored a low compared to high number of points, $F(1, 317) = 11.24$, $p = .001$. In contrast, when a same-attribute reference point was provided, evaluations of the same number of assists did not differ significantly between the low and high number of points conditions, $F(1, 317) = .31$, $p > .57$ (interaction: $F(1, 317) = 7.63$, $p = .006$). In the joint evaluation condition, evaluations of the same number of assists were identical for the two players despite differences in the number of points scored, $F(1, 80) = 0$, $p > .99$.			
Study 3a—Test Score ($N = 101$)			
GPA: 2.1	GMAT: 600		4.90 (1.86)
GPA: 3.7	GMAT: 600		6.92 (1.32)
Main Findings: Participants rated the same GMAT score as significantly better when the candidate's GPA was high compared to low, $t(99) = 6.23$, $p < .001$.			
Study 3b—Hurricane ($N = 141$)			
Category: 2	Economic Impact: $100 million		6.27 (2.04)
Category: 5	Economic Impact: $100 million		6.97 (1.78)
Main Findings: Participants perceived the same amount of economic impact to be significantly larger when the hurricane's category was high compared to low, $t(139) = 2.19$, $p = .03$.			
Study 3c—Electric Guitar ($N = 143$)			
Price: $99	Pickups: Diamond Plus		5.82 (2.02)
Price: $1,199	Pickups: Diamond Plus		6.62 (1.64)
Main Findings: Participants evaluated the same pickups as significantly better when the guitar was relatively more expensive, $t(141) = 2.60$, $p = .01$.			
Study 4–Test Score ($N = 174$)			
GPA: 2.2	ACT: 7		5.53 (1.50)
GPA: 3.6	ACT: 7		6.67 (1.14)
GPA: 2.2	ACT: 7	Joint Evaluation	6.44 (1.61)
GPA: 3.6	ACT: 7		6.86 (1.31)

(*table continues*)

Table 1 (*continued*)

Reference	Target	Moderator	Mean (*SD*)
Main Findings: In the separate evaluation conditions, an identical ACT score was seen as better when the candidate had high compared to low GPA, $t(113) = 4.55$, $p < .001$. The effect of the reference value was largely diminished (but still statistically significant) in magnitude in the joint evaluation condition, $t(58) = 2.14$, $p = .037$. Mediation analyses corroborated that the effect of our GPA manipulation on people's evaluation of the ACT score was mediated by the evaluation of the candidate's GPA score.			
Study 5–Academic vs. Athletic Performance ($N = 402$)			
GPA: 2.1	GMAT: 600	Academic	4.71 (1.84)
GPA: 3.7	GMAT: 600		6.40 (1.36)
GPA: 2.1	Discus throw: 45.10m	Athletic	6.00 (1.47)
GPA: 3.7	Discus throw: 45.10m		6.42 (1.37)
Main Findings: Participants evaluated the same GMAT score more positively when the applicant had a high compared to low GPA, $F(1, 398) = 61.79$, $p < .001$. The effect of our GPA manipulation on judgments about identical discus throw records was weaker in comparison, $F(1, 398) = 3.75$, $p = .053$ (interaction: $F(1, 398) = 17.54$, $p < .001$).			
Study 6–TV ($N = 405$)			
24“	16 pounds	High Evaluability	4.66 (2.19)
65”	16 pounds	(Pounds)	3.65 (1.98)
24“	16 tiris	Low Evaluability	4.27 (1.77)
65”	16 tiris	(Tiris)	6.02 (1.71)
Main Findings: When weight was expressed in pounds, the TV was perceived to be heavier when its screen was small ($M = 4.66$, $SD = 2.19$) compared to large ($M = 3.65$, $SD = 1.98$), $F(1, 401) = 14.00$, $p < .001$. In contrast, when weight was expressed in tiris, the TV was perceived to be heavier when its screen was large ($M = 6.02$, $SD = 1.71$) as opposed to small, ($M = 4.27$, $SD = 1.77$), $F(1, 401) = 42.15$, $p < .001$.			
Study 7–Test Score ($N = 404$)			
SAT: 10%	GMAT: 34%	High Compatibility	4.58 (1.91)
SAT: 99%	GMAT: 34%	(Percentile)	3.83 (1.94)
SAT: 1100/2400	GMAT: 34%	Low Compatibility	3.20 (1.41)
SAT: 2300/2400	GMAT: 34%	(Absolute)	5.34 (2.26)
Main Findings: When the two attributes were expressed in the same scale (i.e., percentile SAT conditions), the same GMAT was perceived to be higher when the applicant's SAT was low compared to high, $F(1, 400) = 7.67$, $p = .006$. In contrast, when the two attributes were expressed in different scales, (i.e., absolute SAT conditions), the same GMAT was perceived to be higher when the applicant had a high compared to low SAT, $F(1, 400) = 64.37$, $p < .001$ (interaction: $F(1, 400) = 58.09$, $p < .001$.			
Study 8–NBA ($N = 305$)			
Points: 4	Efficiency: 5	High Compatibility	48.66 (24.25)
Points: 32	Efficiency: 5	(Integer)	41.27 (30.25)
Points: 4	Efficiency: 5.23	Low Compatibility	34.99 (21.71)
Points: 32	Efficiency: 5.23	(Decimal)	51.77 (25.34)
Main Findings: In the integer target conditions, the same efficiency score was perceived to be higher when the player had scored a low compared to high number of points, $F(1, 301) = 3.21$, $p = .074$. In the decimal target conditions, the same efficiency score was perceived to be higher when the player had scored a high compared to low number of points, $F(1, 301) = 16.24$, $p < .001$ (interaction: $F(1, 301) = 17.01$, $p < .001$).			
Study 9–Hotel ($N = 405$)			
Comfort: 6	Cleanliness: 6	High Compatibility	5.92 (1.64)
Comfort: 10	Cleanliness: 6	(Numeric)	4.63 (1.48)
Comfort: Very poor	Cleanliness: 6	Low Compatibility	4.74 (1.60)
Comfort: Superb	Cleanliness: 6	(Verbal)	5.20 (1.76)
Main Findings: When the two attributes were expressed in the same scale (i.e., numeric comfort conditions), the same cleanliness was perceived to be higher when the hotel's comfort was low compared to high, $F(1, 401) = 32.02$, $p < .001$. In contrast, when the two attributes were expressed in different scales, (i.e., verbal comfort conditions), the same cleanliness was perceived to be higher when the hotel's comfort was high compared to low, $F(1, 401) = 3.96$, $p = .047$ (interaction: $F(1, 401) = 29.29$, $p < .001$).			
Study 10–Test Score ($N = 419$)			
		Target Evaluation	Overall Preference
SAT: 10%	GMAT: 34%	4.72 (2.09)	3.87 (1.82)
SAT: 99%	GMAT: 34%	3.98 (1.95)	4.30 (1.45)
SAT: 1100/2400	GMAT: 34%	3.69 (1.76)	2.83 (1.23)

(*table continues*)

Table 1 (*continued*)

Reference	Target	Moderator	Mean (*SD*)
SAT: 2300/2400	GMAT: 34%	4.58 (1.93)	4.70 (1.49)

Main Findings—Target Evaluation: When the two attributes were expressed in the same scale (i.e., percentile SAT conditions), the same GMAT was perceived to be higher when the applicant's SAT was low compared to high, $F(1, 415) = 7.63$, $p = .006$. In contrast, when the two attributes were expressed in different scales (i.e., absolute SAT conditions), the same GMAT was perceived to be higher when the applicant had a high compared to low SAT, $F(1, 415) = 11.06$, $p = .001$ (interaction: $F(1, 415) = 18.52$, $p < .001$).

Main Findings—Overall Preference: In the absolute SAT conditions, overall preference for the applicant was much higher when the applicant had a high compared to low SAT, $F(1, 415) = 81.03$, $p < .001$. This effect was significantly attenuated—but not eliminated—in the percentile SAT conditions, $F(1, 415) = 4.37$, $p = .037$ (interaction: $F(1, 415) = 23.61$, $p < .001$).

Main Findings—Mediation Analyses: When the two attributes are expressed in different scales (thus being relatively harder to compare), participants evaluate the GMAT score as better when the applicant has a high (vs. low) SAT, and, in turn, rate the high-SAT applicant more positively. When the two scores are expressed in the same scale (thus being relatively easier to compare), participants evaluate the GMAT score as better when the applicant has a low (vs. high) SAT and, in turn, rate the high-SAT applicant less positively.

earthquake was 50 in both conditions. Participants were asked, "In your opinion, how large is the number of people left homeless by the earthquake?" (1= *very small* to 9 = *very large*). The same number (i.e., 50 people) was perceived to be larger when the number of fatalities was low ($M = 4.64$, $SD = 2.21$) compared to high ($M = 3.39$, $SD = 2.20$), $t(199) = 4.03$, $p < .001$.

Study 2

We previously argued that the impact of another (reference) attribute on evaluations of a target attribute should be stronger when there is lack of information about other values on the target attribute. Thus, we hypothesize that individuals will be more likely to evaluate the target without relying on information about other attributes when provided with information about other values on the target attribute. Study 2 tests this hypothesis.

Method

To test this hypothesis, participants in Study 2 ($N = 402$: 51% female; $M_{age} = 35.72$ years, MTurk) were randomly assigned to one of five conditions of a variation of Study 1a. Four conditions involved separate evaluations, while the fifth condition involved joint evaluations. All participants were asked to imagine that they were watching an NBA game. Participants in the separate evaluation conditions were randomly assigned to one of four conditions of a 2 (number of points: low vs. high) by 2 (reference point: no vs. yes) design. Participants were presented with a table featuring information about a player's stats (labeled "Player A"). In the low (high) points conditions the player had scored 5 (33) points and he had handed out 4 assists. Participants in the "no reference point" conditions did not receive any additional information. Participants in the "reference point" conditions were additionally told that "an average NBA player hands out 1.6 assists per game." This information was derived from actual stats about the players at the time of the study (https://www.foxsports.com/nba/) and corresponded to the average number of assists per game by players with an average ranking. Participants in the joint evaluation condition were presented with a table featuring information about both players' stats (labeled "Player A" and "Player B") without the reference point manipulation. Similar to Study 1a, participants were asked to evaluate the number of assists handed out by the player(s) (slider bar: 0= *very low* to 100 = *very high*). We expected that our effect would be attenuated in the separate evaluation reference point conditions and in the joint evaluation condition because same-attribute information is generally more diagnostic than information from another (reference) attribute.

Results

We found support for our predictions. In separate evaluation, we found a significant two-way interaction (see Table 1) between our number of points and reference point manipulations on assists evaluations, $F(1, 317) = 7.63$, $p = .006$. Replicating the result of Study 1a, when no same-attribute reference point was provided, the same number of assists was perceived to be higher when the player had scored a low ($M = 43.71$, $SD = 20.63$) compared to high number of points ($M = 31.06$, $SD = 22.46$), $F(1, 317) = 11.24$, $p = .001$. In contrast, when a same-attribute reference point was provided, evaluations of the same number of assists did not differ significantly between the low ($M = 53.43$, $SD = 24.51$) and high number of points conditions ($M = 55.54$, $SD = 27.61$), $F(1, 317) = 0.31$, $p > .57$. Similarly, in the joint evaluation condition, evaluations of the same number of assists were identical for the two players despite differences in the number of points scored (low number of points: $M = 39.51$, $SD = 23.07$; high number of points: $M = 39.52$, $SD = 23.07$), $F(1, 80) = 0$, $p > .99$. Therefore, Study 2 lends support to our proposition that the impact of a reference attribute on evaluations of the target attribute diminishes when same-attribute information is made available.

Studies 3a–3c

Participants in Studies 3a–3c were randomly assigned to one of two conditions. Participants in both conditions were asked to evaluate the same (target) attribute value of a given stimulus, while we manipulated the level of a second (reference) attribute. Consistent with an interattribute inference process, we find robust evidence that identical target values are perceived to be larger when reference values are high as opposed to low (see Table 1).

Study 3a

Participants in Study 3a ($N = 101$: 51% female; $M_{age} = 33.5$ years, MTurk) were asked to imagine that they were the hiring manager of a company and that they were considering a candidate for an entry-level job. In the low (high) GPA condition, the candidate had a college GPA of 2.1 (3.7) and a GMAT score of

600. Information was presented in table format with a different column for each attribute. We asked participants, "In your opinion, how good is the candidate's GMAT score?" (1= *very bad* to 9 = *very good*). Participants rated the same GMAT score of 600 as significantly better when the candidate's GPA was high (M = 6.92, SD = 1.32) compared to low (M = 4.90, SD = 1.86), $t(99)$ = 6.23, $p < .001$.

Study 3b

Participants in Study 3b (N = 141: 54% female; M_{age} = 34.7 years, MTurk) were asked to imagine that a hurricane had hit the U.S. In the low (high) category condition, the hurricane's category was 2 (5) and its economic impact was $100 million. Information was presented in table format with a different column for each attribute. We asked participants, "In your opinion, how big is the hurricane's economic impact?" (1= *very small* to 9 = *very big*). Participants perceived the same amount of economic impact (i.e., $100 million) to be significantly larger when the hurricane's category was high (M = 6.97, SD = 1.78) compared to low (M = 6.27, SD = 2.04), $t(139)$ = 2.19, p = .03.

Study 3c

Participants in Study 3c (N = 143: 57% male; M_{age} = 34 years, MTurk) were asked to imagine that they were considering buying an electric guitar. In the low (high) price condition, the guitar was priced at $99 ($1,199) and featured Diamond Plus pickups. Information was based on actual electric guitars sold by Schecter and was presented in table format with a different column for each attribute. We asked participants, "In your opinion, how good are the guitar's pickups?" (1= *very bad* to 9 = *very good*). Participants evaluated the same pickups as significantly better when the guitar was relatively more expensive (M = 6.62, SD = 1.64 vs. M = 5.82, SD = 2.02), $t(141)$ = 2.60, p = .01. Note that this result is different from standard price-quality inference effects for two reasons. First, whereas responses elicited in price-quality research pertain to the overall quality of the product, in Study 3c we measure perceived performance on a given attribute rather than overall quality. Second and most important, in Study 3c, the level of the quality attribute is stated, whereas in price-quality research no information about the quality attribute is provided at all. Participants in those studies are simply asked to make inferences about missing (overall quality) information.

Study 4

Study 4 serves two goals. First, we seek to provide direct evidence for our hypothesis that judgments about the attractiveness of the reference value correlate positively with evaluations of the (less evaluable) target value consistent with the interattribute inference process. Second, we examine whether the impact of another (reference) attribute on evaluations of the target during the interattribute inference process is attenuated in joint compared to separate evaluation mode, because same-attribute information is more diagnostic than information from another attribute.

Method

Participants in Study 4 (N = 174: 51% male; M_{age} = 35.47 years, MTurk) were randomly assigned to one of three conditions. All participants were asked to imagine that they were the hiring manager of a company. Participants in the low (high) separate GPA conditions were asked to imagine that they were considering a candidate for an entry-level job. In the low (high) separate GPA condition, the candidate had a college GPA of 2.2 (3.6) and an Amos Creativity Test (ACT) score of 7.[1] Participants in the joint condition were exposed to both candidates at the same time. Participants were asked to rate the score of the candidate(s) on the Amos Creativity Test (1= *very bad* to 9 = *very good*). We further asked participants to evaluate the candidate's college GPA using the same 9-point scale in order to examine whether evaluations of performance on one attribute (i.e., GPA) influence evaluations of performance on another dimension (i.e., ACT). The order of the two measures was counterbalanced and did not moderate our effects.

Results

We first replicated results of Studies 3a–3c in the separate evaluation conditions (see Table 1). An identical score (i.e., 7) on the ACT was seen as better when the candidate had high (M = 6.67, SD = 1.14) compared to low GPA (M = 5.53, SD = 1.50), $t(113)$ = 4.55, $p < .001$. In addition, as we would expect, a high GPA score was rated as better (M = 7.30, SD = 1.07) compared to a low GPA score (M = 3.41, SD = 1.74), $t(113)$ = 14.41, $p <$ .001. Importantly, a mediation analysis (Hayes, 2013) corroborated that the effect of our GPA manipulation on people's evaluation of the ACT score was mediated by the evaluation of the candidate's GPA score (10,000 bootstraps; 95% bias-corrected confidence intervals; ab = 0.969, SE = 0.327, *95% LLCI* = 0.358, *ULCI* = 1.663). This result provides support for our proposition that, during the interattribute inference process, participants' evaluations of a given target value are contingent on the perceived intensity of the reference value. Increasing the candidate's GPA score strengthened subjective evaluations of his or her GPA, which, in turn, had a positive spillover effect on evaluations of the target attribute ACT. In other words, people evaluated a candidate's ACT score in a manner that was consistent with their beliefs about his or her GPA score. Identical ACT scores were rated more positively the more participants felt that the candidate's GPA score was good.

The effect of the reference value (i.e., high vs. low GPA) was largely diminished (but still statistically significant) in magnitude in the joint evaluation condition. Despite the fact that actual scores were the same (both scored a 7 on the ACT), participants rated the low GPA candidate's ACT score as worse compared to that of the high GPA candidate (M = 6.44, SD = 1.61 vs. M = 6.86, SD = 1.31), $t(58)$ = 2.14, p = .037. That is, the difference in ACT score evaluations between participants with high versus low GPA was reduced by 63% when joint evaluation allowed a within-attribute-process inference that the two people scored the same on the ACT. (Note that because the results in the separate conditions were between-participants whereas this result is within-participant, no interaction effect could be computed.) As one would expect, participants also rated the low GPA candidate's GPA score as worse compared to that of the high GPA candidate (M = 3.66, SD = 1.68 vs. M = 7.81, SD = 0.90), $t(58)$ = 17.19, $p < .001$.

[1] ACT is a fictitious test that we named in memory of Amos Tversky.

Finally, based on guidelines by Judd, Kenny, and McClelland (2001), we regressed the difference in evaluations of the candidates' ACT scores on the difference and on the sum (mean-centered) of the evaluations of the candidates' GPA and found that the former predictor was marginally significant ($B = 0.234$, $SE = 0.128$, $p = .072$); this result suggests that whereas the effect of GPA is greatly diminished, evaluations of the candidates' GPA scores mediated the effect of our manipulation on evaluations of the candidates' ACT scores. In conclusion, we found support for the proposed interattribute inference process, which postulates that people evaluate the attractiveness of a target attribute value based on the reference value. We further found evidence that the impact of the reference value on evaluations of the target diminishes when more diagnostic inputs (i.e., same-attribute information) are made available.

Study 5

In Study 5 we test our hypothesis that the extent to which the two attributes are seen as correlated moderates the impact of the reference value on evaluations during the interattribute inference process. We expect that inferences about the target attribute value are more likely to be influenced by a reference value when the two attributes are perceived to be highly correlated.

Method

Participants in Study 5 ($N = 402$: 57% male; $M_{\text{age}} = 33.07$ years, MTurk) were randomly assigned to one of four conditions of a 2 (GPA: low vs. high) by 2 (DV: GMAT vs. discus) between-participants design. All participants were asked to imagine that they were the admissions manager at a local university and that they were considering the following applicant. In the low (high) GPA condition, the applicant had a high school GPA of 2.1 (3.7). In the GMAT conditions, the applicant had a GMAT score of 600. In the discus throw conditions, the applicant had a discus throw personal record of 45.10 m. Information was presented in table format with a different column for each attribute (High School GPA and GMAT Score or High School GPA and Discus Throw Record). Participants were asked, "In your opinion, how good is the candidate's GMAT score (discus throw record)?" (1= *very bad* to 9 = *very good*).

On the next page, as a manipulation check, we measured participants' perceived interattribute correlation by asking them a pair of questions: (1) "In your opinion, how diagnostic is one's high school GPA of his or her GMAT score (discus throw record)?" (1= *not diagnostic at all* to 9 = *very diagnostic*) and (2) "In your opinion, to what extent is one's high school GPA correlated with his or her GMAT score (discus throw record)?" (1= *very weakly correlated* to 9 = *very strongly correlated*; we averaged the two items to create an index of relatedness, $r = .81$). Participants perceived one's high school GPA to be more strongly related to his or her GMAT score ($M = 5.61$, $SD = 1.52$) compared to his or her discus throw record ($M = 2.48$, $SD = 1.78$), $F(1, 398) = 359.47$, $p < .001$. We also measured participants' personal knowledge about the different attributes (e.g., "How much do you know about the GMAT test (discus throw)?" 1= *I do not know much about it* to 9 = *I know a lot about it*). Participants' knowledge about both the GMAT test ($M = 2.91$, $SD = 2.18$) and discus throw ($M = 2.45$, $SD = 1.87$) was quite limited (significantly below the scale midpoint, p values <0.001).

Results

We found a significant two-way interaction (see Table 1) between our GPA manipulation and the dependent variable that participants were asked to evaluate, $F(1, 398) = 17.54$, $p < .001$. The effect of our GPA manipulation on judgments about identical GMAT scores was strong and significant, $F(1, 398) = 61.79$, $p < .001$. Participants evaluated the same GMAT score more positively when the applicant had a high ($M = 6.40$, $SD = 1.36$) compared to low GPA ($M = 4.71$, $SD = 1.84$). The effect of our GPA manipulation on judgments about identical discus throw records was weaker in comparison, yet still statistically significant, $F(1, 398) = 3.75$, $p = .053$. Participants evaluated the same discus throw record more positively when the applicant had a high ($M = 6.42$, $SD = 1.37$) compared to low GPA ($M = 6.00$, $SD = 1.47$). Therefore, consistent with our prediction, we find that the impact of the reference value on evaluations during interattribute inferences is contingent on the degree to which the two attributes are seen as correlated. When perceived interattribute correlation is high, such as between GPA and GMAT, the reference attribute exerts a stronger influence on judgments about a target attribute value.

Study 6

We have argued that the interattribute inference process should be stronger as the target attribute becomes less evaluable. Conversely, the interattribute comparison process should be stronger as the target attribute becomes more evaluable. We test our propositions in Study 6 by manipulating the evaluability of the target attribute. The context of Study 6 was televisions, a product category with which most people are relatively familiar. The reference attribute was screen size (expressed in inches). The target attribute was weight. In the high evaluability conditions, weight was expressed in pounds (i.e., the default unit for our sample of U.S. participants). In the low evaluability conditions, weight was expressed in a fictitious unit labeled "tiris." By using a fictitious unit, we were able to keep the target value constant across conditions.

Participants in Study 6 ($N = 405$: 61% female; $M_{\text{age}} = 36.5$ years, MTurk) were presented with a table featuring information about a TV (labeled "TV A"). They were randomly assigned to one of four conditions of a 2 (screen size: small vs. large) by 2 (weight: pounds vs. tiris) between-participants design. In the small (big) screen size conditions, the TV's screen size was 24" (65"). In the pounds (tiris) conditions, the TV's weight was 16 pounds (tiris). Participants were asked, "In your opinion, how heavy is TV A?" (1= *not at all heavy* to 9 = *very heavy*). After evaluating the TV's weight, we administered our manipulation check. Participants were asked to indicate how easy it was to evaluate the TV's weight (scale: 1= *very difficult* to 9 = *very easy*). Participants indicated that it was easier to evaluate the TV's weight when it was expressed in pounds as opposed to tiris ($M_{pounds} = 7.11$, $SD_{\text{pounds}} = 1.81$ vs. $M_{tiris} = 4.31$, $SD_{\text{tiris}} = 2.64$), $F(1, 401) = 156.08$, $p < .001$; this effect was not moderated by our screen size manipulation ($p > .16$). Therefore, we expected that participants would be more likely to evaluate the target attribute weight based on an interattribute comparison process when weight was expressed in pounds. In contrast, we expected that participants would be more likely to evaluate the target based on an interattribute inference process when weight was expressed in tiris.

We found an interaction between our screen size and weight factors, $F(1, 401) = 52.27$, $p < .001$ (see Table 1). Consistent with the interattribute comparison process, when weight was expressed in pounds, the TV was perceived to be heavier when its screen was small ($M = 4.66$, $SD = 2.19$) compared to large ($M = 3.65$, $SD = 1.98$), $F(1, 401) = 14.00$, $p < .001$. In contrast, consistent with the interattribute inference process, when weight was expressed in tiris, the TV was perceived to be heavier when its screen was large ($M = 6.02$, $SD = 1.71$) as opposed to small, ($M = 4.27$, $SD = 1.77$), $F(1, 401) = 42.15$, $p < .001$. In conclusion, the data of Study 6 suggest that the direction of the effect of the reference value on evaluations of the target value is contingent on the evaluability of the target attribute.

Study 7

In Study 7, we examine whether attribute value compatibility influences the process through which individuals evaluate the target. Holding our target attribute constant, we manipulate whether the reference attribute is expressed in the same (vs. different) scale as the target. We predicted that participants would be more likely to evaluate the target based on a comparison rather than an inference process when both attributes values were compatible, because they were expressed in the same scale (attribute value compatibility should make it easier to compare the two values). Conversely, we predicted that participants would be more likely to evaluate the target based on an interattribute inference rather than a comparison process when the two attributes are expressed in different scales, as this reduction in attribute value compatibility should make it more difficult to compare the two values.

Participants in Study 7 ($N = 404$: 58% female; $M_{\text{age}} = 35.6$ years, MTurk) were asked to imagine that they were the admissions manager at a local university and that they were considering the following applicant. Participants were presented with a table featuring information about the applicant's (labeled "Applicant A") SAT and GMAT score. Participants in all conditions were told that the applicant's GMAT percentile was 34%. Participants were randomly assigned to either a same-scale (i.e., percentile SAT) or different-scale (i.e., absolute SAT) reference attribute condition. In the low (high) absolute SAT condition, the applicant had an SAT of 1100 (2300) out of 2400. In the low (high) percentile SAT condition, the applicant's SAT percentile was 10% (99%); we converted absolute SAT scores to the corresponding percentiles based on information provided by PrepScholar (https://blog.prepscholar.com/sat-percentiles-high-precision-2016). Participants were asked to evaluate the applicant's GMAT (1= *very bad* to 9 = *very good*). To ascertain that any effects would really be driven by attribute value compatibility and not by reference attribute evaluability (note that target evaluability was already controlled for because the target attribute format was held constant in this study), we ran a separate study. Participants in this separate study ($N = 50$: 56% male; $M_{\text{age}} = 36.3$ years, MTurk) were provided with both absolute and percentile SAT scores and indicated that the two are equally easy to evaluate (scale: 1= *very difficult* to 9 = *very easy*; $M_{absolute} = 7.42$, $SD_{\text{absolute}} = 1.89$ vs. $M_{percentile} = 7.44$, $SD_{\text{percentile}} = 2.35$, $t(49) = 0.06$, $p > .95$). Therefore, predicted differences in the evaluation process that individuals rely on to evaluate the target should be predicated on the compatibility of the two values, rather than differences in the evaluability of the reference attribute.

Going back to the main study, we found the expected interaction between our SAT scale and SAT value factors, $F(1, 400) = 58.09$, $p < .001$ (see Table 1). When the two attributes were expressed in the same scale (i.e., percentile SAT conditions), the same GMAT (i.e., 34%) was perceived to be higher when the applicant's SAT was low ($M = 4.58$, $SD = 1.91$) compared to high ($M = 3.83$, $SD = 1.94$), $F(1, 400) = 7.67$, $p = .006$. In contrast, when the two attributes were expressed in different scales, (i.e., absolute SAT conditions), the same GMAT (i.e., 34%) was perceived to be higher when the applicant had a high ($M = 5.34$, $SD = 2.26$) compared to low SAT ($M = 3.20$, $SD = 1.41$), $F(1, 400) = 64.37$, $p < .001$. Therefore, Study 7 provides support to our proposition that the direction of the effect of the reference value on evaluations of the target value is contingent on the compatibility of the two values.

Study 8

In Study 7 we showed that evaluators are more likely to employ a comparison process when attribute value compatibility is relatively high, such as when both attributes are measured on the same scale. In our theoretical development, we argued that attribute value compatibility may also be contingent on the format of the two values. For example, attribute value compatibility should be higher when both attributes assume integer values as opposed to when one attribute assumes an integer value while the other assumes a decimal value. We tested this prediction in Study 8. Participants ($N = 305$: 53% female; $M_{\text{age}} = 35.8$ years, MTurk) were asked to imagine that they were watching an NBA game. Participants were presented with a table featuring information about a player's stats (labeled "Player A"). In the low (high) points conditions the player had scored 4 (32) points. In the integer (decimal) target conditions the player had 5 (5.23) efficiency. Participants were asked to evaluate the efficiency of Player A (slider bar: 0= *very low* to 100 = *very high*). To ascertain that any effects would really be driven by attribute value compatibility and not by target attribute evaluability (note that reference evaluability was already controlled for because the reference attribute format was held constant in this study), we ran a separate study. Participants in this separate study ($N = 81$: 56% female; $M_{\text{age}} = 38.1$ years, MTurk) were provided with either the decimal or the integer efficiency and were asked to rate how easy it is to evaluate the target (scale: 1= *very difficult* to 9 = *very easy*). There were no significant differences in evaluability ratings ($M_{integer} = 3.73$, $SD_{\text{integer}} = 2.24$ vs. $M_{decimal} = 4.17$, $SD_{\text{decimal}} = 2.64$, $t(79) = 0.82$, $p > .41$). Therefore, predicted differences in the evaluation process that individuals rely on to evaluate the target should be predicated on the compatibility of the two values, rather than differences in the evaluability of the target attribute. We expected that participants in the integer target conditions would be more likely to evaluate the target by comparing it with the reference value. Conversely, we expected that participants in the decimal target conditions would be more likely to evaluate the target by making an inference based on the reference value.

We found an interaction between the number of points and the format of the target attribute, $F(1, 301) = 17.01$, $p < .001$ (see Table 1). Consistent with our predictions, in the integer target conditions, the same efficiency score (i.e., 5) was perceived to be higher when the player had scored a low ($M = 48.66$, $SD = 24.25$) compared to high number of points ($M = 41.27$, $SD = 30.25$), $F(1, 301) = 3.21$, $p = .074$. In sharp contrast, in the decimal target conditions, the same efficiency score (i.e., 5.23) was perceived to be higher when the player had scored a high ($M = 51.77$, $SD = 25.34$) compared to low number of points ($M = 34.99$, $SD = 21.71$), $F(1, 301) = 16.24$, $p < .001$. Therefore, the data of Study 8 suggest that the direction of the effect of the reference value on evaluations of the target value is contingent on attribute value compatibility. When the two values are more compatible (i.e., when both values are integers), we find that the target value is perceived to be smaller when the reference value is high as opposed to low. In contrast, when the two values are less compatible (i.e., when the target value is decimal while the reference value is integer), the target value is perceived to be larger when the reference value is high compared to low.

Study 9

In our previous two studies, we showed that evaluators are more likely to rely on an interattribute comparison process over an interattribute inference process when attribute value compatibility is high as opposed to low. In this study, we employ an additional manipulation of the format of the two values that should influence attribute value compatibility. In particular, attribute value compatibility should be higher when both attributes assume numeric values compared to when one of the two attributes assumes a verbal (or categorical) value. We tested this prediction in Study 9. Participants in Study 9 ($N = 405$: 57% female; $M_{\text{age}} = 36.3$ years, MTurk) were asked to imagine that they were browsing hotels at Booking.com, and that they were reading reviews about the following hotel. Participants were presented with a table featuring information about the hotel's comfort and cleanliness ratings. Participants in all conditions were told that the hotel's cleanliness was 6. Participants were randomly assigned to either a same-scale (i.e., numeric comfort) or different-scale (i.e., verbal comfort) reference attribute condition. In the low (high) numeric comfort condition, the hotel had a comfort rating of 2 (10). In the low (high) verbal comfort condition, the hotel had a comfort rating of "very poor" ("superb"); we converted numeric ratings to the corresponding verbal ratings based on information provided by Booking.com. Participants were asked to evaluate the hotel's cleanliness (1= *not at all clean* to 9 = *very clean*). To ascertain that any effects would really be driven by attribute value compatibility and not by reference attribute evaluability (note that target evaluability was already controlled for because the target attribute format was held constant in this study), we ran a separate study. Participants in this separate study ($N = 52$: 54% female; $M_{\text{age}} = 34.9$ years, MTurk) were provided with both numeric and verbal comfort scores and indicated that one is not significantly easier to evaluate than the other (scale: 1= *very difficult* to 9 = *very easy*; $M_{numeric} = 7.71$, $SD_{\text{numeric}} = 1.80$ vs. $M_{verbal} = 8.00$, $SD_{\text{verbal}} = 1.44$, $t(51) = 1.18$, $p > .26$). Therefore, similar to prior studies, predicted differences in the evaluation process that individuals rely on to evaluate the target should be predicated on the compatibility of the two values, rather than differences in the evaluability of the reference attribute.

Going back to the main study, we found a significant interaction between our comfort scale and comfort value factors, $F(1, 401) = 29.29$, $p < .001$ (see Table 1). When the two attributes were expressed in the same scale (i.e., numeric comfort conditions), the same cleanliness (i.e., 6) was perceived to be higher when the hotel's comfort was low ($M = 5.92$, $SD = 1.64$) compared to high ($M = 4.63$, $SD = 1.48$), $F(1, 401) = 32.02$, $p < .001$. In contrast, when the two attributes were expressed in different scales (i.e., verbal comfort conditions), the same cleanliness (i.e., 6) was perceived to be higher when the hotel's comfort was high ($M = 5.20$, $SD = 1.76$) compared to low ($M = 4.74$, $SD = 1.60$), $F(1, 401) = 3.96$, $p = .047$. Therefore, Study 9 provides further support for our proposition that the direction of the effect of the reference value on evaluations of the target value is contingent on the compatibility of the two values.

Study 10

In our final study, we examine whether the two evaluation processes affect individuals' overall preference for the target besides their attribute-level evaluations. Participants in Study 10 ($N = 419$: 51% male; $M_{\text{age}} = 34.3$ years, MTurk) were randomly assigned to one of the four conditions of Study 7. Similar to Study 7, participants were asked to evaluate the applicant's GMAT (1= *very bad* to 9 = *very good*). Participants were additionally asked to express their overall preference for the candidate on two items: (1) "the candidate should be admitted to the graduate program" and (2) "the candidate is a good applicant" (1= *strongly disagree* to 7 = *strongly agree*; we combined the two items to create an index of preference, $r = .89$). We randomized the order of our measures.

First, we replicated results of Study 7. We found an interaction between our SAT scale and SAT value factors on GMAT evaluations, $F(1, 415) = 18.52$, $p < .001$ (see Table 1). When the two attributes were expressed in the same scale (i.e., percentile SAT conditions), the same GMAT (i.e., 34%) was perceived to be higher when the applicant's SAT was low ($M = 4.72$, $SD = 2.09$) compared to high ($M = 3.98$, $SD = 1.95$), $F(1, 415) = 7.63$, $p = .006$. In contrast, when the two attributes were expressed in different scales, (i.e., absolute SAT conditions), the same GMAT (i.e., 34%) was perceived to be higher when the applicant had a high ($M = 4.58$, $SD = 1.93$) compared to low SAT ($M = 3.69$, $SD = 1.76$), $F(1, 415) = 11.06$, $p = .001$.

Importantly, we found an interaction between our two factors on overall evaluations of the target, $F(1, 415) = 23.61$, $p < .001$ (see Table 1). In the absolute SAT conditions, overall preference for the applicant was much higher when the applicant had a high ($M = 4.70$, $SD = 1.49$) compared to low SAT ($M = 2.83$, $SD = 1.23$), $F(1, 415) = 81.03$, $p < .001$. This effect was significantly attenuated—but not eliminated—in the percentile SAT conditions. Overall preference for the applicant was slightly higher when the applicant had a high ($M = 4.30$, $SD = 1.45$) compared to low SAT ($M = 3.87$, $SD = 1.82$), $F(1, 415) = 4.37$, $p = .037$. The fact that participants in the percentile SAT conditions still preferred—yet to a lesser extent—the high-SAT applicant over the low-SAT applicant unsurprisingly implies that the importance weight afforded to the presumably much more evaluable SAT is larger than the weight afforded to the presumably less evaluable GMAT. To recap, holding GMAT scores constant, participants expressed a stronger preference for the applicant with the high compared to low SAT score. However, the magnitude of the effect of the SAT score on preference was contingent on whether the two attributes were measured on the same scale. The mean difference in preference for the two applicants was larger when the two attributes were expressed in different scales (i.e., 1.87 points) compared to when they were expressed in the same scale (i.e., 0.43 points). We surmise that this result is observed because the same GMAT score is evaluated differently depending on whether it can be easily compared to the SAT score. When the two attributes are expressed in different scales (thus being relatively harder to compare), participants evaluate the GMAT score as better when the applicant has a high (vs. low) SAT and, in turn, rate the high-SAT applicant more positively. A mediation analysis provides support for this proposition (10,000 bootstraps; 95% bias-corrected confidence intervals; $ab = 0.471$, $SE = 0.141$, $95\%\ LLCI = 0.200$, $ULCI = 0.747$). In contrast, when the two scores are expressed in the same scale (thus being relatively easier to compare), participants evaluate the GMAT score as better when the applicant has a low (vs. high) SAT and, in turn, rate the high-SAT applicant less positively. A mediation analysis provides support for this proposition (10,000 bootstraps; 95% bias-corrected confidence intervals; $ab = -0.394$, $SE = 0.156$, $95\%\ LLCI = -0.709$, $ULCI = -0.093$). The difference between the two indirect effects was statistically significant (*index of moderated mediation* $= -0.864$, $SE = 0.219$, $95\%\ LLCI = -1.300$, $ULCI = -0.445$; Hayes, 2013).

General Discussion

We have advanced a theory that describes how people evaluate attribute values. We propose that attribute evaluations typically involve a target and a reference value. The reference value may be another value on the same attribute, but also a value on another attribute. The evaluator is more likely to rely on a reference value associated with another attribute when there is no information about another value on the target attribute. In the absence of same-attribute information, the individual may make an evaluation about the target value in one of two ways. First, she may compare the target attribute value to the stimulus's value on a reference attribute (i.e., interattribute comparison process). The evaluator is more likely to engage in an interattribute comparison when the target attribute is relatively evaluable and the two attribute values are relatively compatible. Second, she may infer the magnitude of the target value based on her judgment about the extremity (e.g., the goodness or badness) of the stimulus's value on a reference attribute and the perceived correlation between the target attribute and the reference attribute (i.e., interattribute inference process). The evaluator is more likely to make an inference about the target based on the reference when it is relatively hard to evaluate the target and the two attribute values are less compatible.

We found robust evidence for our framework in 14 studies (see Table 1). Because prior research provides ample evidence for evaluations based on within-attribute comparisons (for a review see Hsee et al., 1999), our investigation was focused on providing evidence for the two interattribute evaluation processes. Using compatible attribute values, Studies 1a–1c provide evidence for the interattribute comparison process by showing that, holding target values fixed, targets are perceived to be smaller (or worse) as reference values increase. Study 2 shows that this effect is moderated when same-attribute information is made available. Using incompatible attribute values, Studies 3a–3c provide evidence for the interattribute inference process by showing that, holding target values fixed, targets are perceived to be larger (or better) as reference values increase. Study 4 provides evidence for our theoretical account by showing that evaluations of the reference value correlate positively with evaluations of the target value and by showing that the impact of the inference process on evaluations is attenuated when same-attribute information is made available. Furthermore, Study 5 presents evidence that the perceived correlation between the reference and the target moderates the strength of the inference process. In Studies 6–9 we provide evidence for our theoretical framework by showing that the effect of the reference value on evaluations of the target is contingent on the target's evaluability and on attribute value compatibility. Finally, in Study 10 we provide evidence that the effect of the reference value on evaluations of a target value extends to evaluators' overall preference for the target.

Theoretical Contribution

Evaluability Theory. Evaluability Theory (Hsee et al., 1999; Hsee & Zhang, 2010)—the most prominent framework in psychology with respect to attribute-level evaluations—discusses factors that determine the evaluability of a given attribute. These factors include the evaluator's knowledge about the target attribute, the nature of the target attribute (i.e., whether the attribute is inherently evaluable), as well as the evaluation mode (joint vs. separate). Nonetheless, Evaluability Theory remains rather silent about whether, when, and how evaluators use information about other (i.e., reference) attributes when evaluating a target. To our knowledge, our theoretical framework is the first to provide an answer to these three questions. We build on Evaluability Theory by adding two evaluation processes whereby individuals rely on information about other (reference) attributes when assessing the magnitude of a target in the absence of same-attribute information. We acknowledge that there may be additional inputs—besides the two processes discussed in our paper—that individuals employ when making attribute-level evaluations. We also acknowledge that there may be other factors—besides target evaluability and attribute value compatibility—that influence the direction and the magnitude of the effect of the reference attribute on the evaluation of the target. Nonetheless, we do believe that our theory and data advance our understanding of how individuals evaluate target attribute values. Our framework is the first that explains how individuals use both within- and between-attribute information to evaluate a target.

Compatibility and alignability in judgments and decisions. Prominent research in psychology argues that an input in a judgment or decision is afforded greater weight as its compatibility with the output increases (Slovic, Griffin, & Tversky, 1990; Tversky, Sattath, & Slovic, 1988; see also Fischer & Hawkins, 1993; Shafir, 1995). This principle is labeled "scale compatibility" (Slovic et al., 1990; Tversky et al., 1988). For example, individuals faced with a task where they need to price different gambles are more likely to focus on payoffs than on probabilities because the former are more compatible with the response scale (i.e., both involve financial amounts). In this article we advance a novel type of compatibility—attribute value compatibility—which captures the extent to which two different inputs—the reference and the target attribute—are compatible in evaluation tasks. When the formatting of the two inputs is more similar—such as when the two attributes are measured on similar scales—attribute value compatibility increases and evaluators are more likely to engage in interattribute comparisons instead of interattribute inferences.

Other work discusses the alignability of different features as a driver of choice (Markman & Loewenstein, 2010; see also dimensional commensurability in Slovic & MacPhillamy, 1974). When a given attribute is alignable, it can be found across different stimuli. In contrast, when an attribute is nonalignable, it may be found in one stimulus but not in the other. For example, camera resolution is an alignable attribute when the decision-maker makes a choice between two smartphones, but it is a nonalignable attribute when the decision-maker chooses between a smartphone and a Bluetooth speaker. Therefore, alignability (or commensurability) captures the extent to which an attribute is common (i.e., alignable) across stimuli. In contrast, attribute value compatibility captures the extent to which two attributes values are compatible (and thus comparable) within a stimulus.

Selective versus comparative processing. According to Sanbonmatsu, Vanous, Hook, Posavac, and Kardes (2011), comparative processing describes the case where decision-makers compare a focal object to other alternatives on a set of attributes, whereas selective processing refers to evaluating the object in isolation without examining other options. Selective option processing is cognitively easier but can lead to less accurate judgments. In our article, we discuss two processes—a comparison and inference process—where both utilize information about two attributes in the context of a single option. In that sense, both processes seem to be two instances of selective processing (because evaluators look at one option). These two processes are moderated when additional options are introduced (in which case most evaluators switch to comparative processing of information). Thus, vis-à-vis Sanbonmatsu et al. (2011), one could argue that our interattribute comparison process is a type of comparative processing between attributes but within a stimulus.

The halo effect. When the target is relatively hard to evaluate and cannot be easily compared to the reference value, we propose and find that evaluations of the target increase as the stimulus's value on the reference attribute increases. To the extent that good performance on the reference attribute boosts general evaluations of the stimulus, this result may seem consistent with the well-known halo effect. According to the halo effect, global evaluations of a stimulus can have an effect on evaluations of its attributes (Nisbett & Wilson, 1977; Thorndike, 1920). For example, in a demonstration of the halo effect by Beckwith and Lehmann (1975), respondents' overall attitude toward a TV show (i.e., general liking or disliking) exerted a strong influence on their attribute-level beliefs—particularly for ambiguous attributes—such as "well-produced or directed." Thus, it may seem possible that our interattribute inference effect (but not our interattribute comparison effect) is mediated by a halo process.

In this article, we posit that a specific reference attribute value can exert a direct influence on the evaluation of another (target) attribute level—an influence that is not mediated by a halo process. To assess the veracity of this postulate, we ran an additional experiment. Specifically, we explored whether our interattribute inference effect would replicate in a situation in which higher levels of the reference attribute would not be accompanied by a more positive overall attitude. We randomly assigned 266 participants (59% female; M_{age} = 36.1 years, MTurk) to one of four conditions of a 2 (overall attitude: less vs. more positive) by 2 (GPA: low vs. high) between-participants design. Participants were presented with information about a student named John. All participants were told that "John goes on an exchange to an Italian university where classes are taught in English but the Italian grading system is used." Participants in the less positive attitude conditions were further told that John "doesn't partake in the exchange experience. Instead of reaching out to Italian students he sits in his room to study whenever he's not in class. When he does have to interact with the Italian students, he is curt and business-like." Participants in the more positive attitude conditions were instead told that John "partakes fully in the exchange experience. He actively reaches out to Italian students and hangs out and studies with them when not in class. When he interacts with the Italian students, he is friendly and warm." Participants were additionally presented with information about John's academic performance. In the low (high) GPA condition, John had a GPA of 2.3 (3.7) and a test score at the Italian university of 28. Information was presented in table format with a different column for each attribute. Participants were asked to indicate their opinion about John (1= *very negative* to 9 = *very positive*). Additionally, participants were asked to rate John's test score at the Italian university (1= *very bad* to 9 = *very good*). All information and questions were presented on the same page. First, our manipulation was successful: participants in the more positive attitude condition evaluated John more positively (M = 8.07, SD = 1.13) compared to their counterparts in the less positive attitude condition (M = 4.17, SD = 1.71), $F(1, 262) = 476.07$, $p < .001$. Our GPA manipulation had no effect on participants' overall attitude toward John, and there was no interaction between our two factors on the same measure (p values $> .80$). Importantly, we replicated our interattribute inference effect. Participants rated the same test score of 28 as significantly better when John's GPA was high (M = 6.98, SD = 1.55) compared to low (M = 3.93, SD = 1.44), $F(1, 262) = 308.02$, $p < .001$. This result was not moderated by our attitude manipulation ($p > .82$). Consistent with a halo effect, we did find that participants rated the same score more positively in the more positive attitude condition (M = 5.92, SD = 2.07) compared to the less positive attitude condition (M = 4.97, SD = 2.11), $F(1, 262) = 29.52$, $p < .001$. Importantly, we found that participants in the "less positive attitude high GPA" condition rated John's test score as significantly better (M = 6.53, SD = 1.50) compared to their counterparts in the "more positive attitude low GPA" condition (M = 4.42, SD = 1.35), $F(1, 262) = 73.42$, $p < .001$. This result is consistent with a direct interattribute inference process but is inconsistent with a halo-mediated process. Our reference attribute manipulation is not just a way to change the overall attitude driving a halo effect.

Assimilation, contrast, and anchoring. Beyond advancing a new theory of attribute-level evaluations, we also provided robust evidence that identical attribute values may receive different evaluations. Our research is not the first to argue that people may evaluate identical attribute values differently. Classic research on assimilation and contrast suggests that people may compare a given attribute value to a salient reference point—such as another value on the same attribute—during evaluation (Sherif & Hovland, 1961; Sherif, Taub, & Hovland, 1958). For instance, an SAT of 1200 may be perceived to be better when evaluated next to an SAT of 600 compared to an SAT of 1000, because 1200 feels much larger in comparison to 600 versus 1000. Bless and Schwarz's (2010) inclusion/exclusion model posits that the same information can lead to assimilation versus contrast effects depending on its use. According to the inclusion/exclusion model, evaluative judgments are based on two mental representations: one of the target and one of the standard or reference point compared to which individuals evaluate the target. Assimilation occurs when a given piece of information is used in forming a representation of the target. Contrast occurs when a given piece of information is used in forming a representation of the comparison standard. With respect to the inclusion/exclusion model, our theory and results imply that whether a reference value is used in forming a representation of the target versus the comparison standard depends on the evaluability of the target and the attribute value compatibility between the target value and the reference value. When the target is hard to evaluate and incompatible with the reference value, the reference value is used in forming the representation of the target. When the target value is relatively evaluable and compatible with the reference value, the reference value is used in forming the representation of the comparison standard.

Relatedly, the anchoring effect suggests that judgments may be biased in the presence of arbitrary anchors such that the value of the judgment is assimilated to the value of an anchor. For instance, participants in a classic study by Tversky and Kahneman (1974) were asked to indicate the percentage of African countries in the United Nations. Participants' estimates were biased in the direction of arbitrary anchors that were determined by a spinning wheel of fortune. Frederick and Mochon (2012) further showed that such anchoring effects can be expected when the measurement scale of the response task is the same as the scale of the reference point (anchor). For example, an anchor (e.g., the weight of a whale) is more likely to have an effect on judgments about a stimulus's weight (e.g., the weight of a raccoon) when the response scale is the same as that of the anchor (i.e., both in pounds) compared to when the scales differ (e.g., pounds vs. tons).

While research on assimilation, contrast, and anchoring focuses on judgments of a (target) attribute value in the presence versus absence of reference points (or anchors) associated with the same attribute of other stimuli or with an arbitrary source such as a wheel of fortune, we provide evidence for similar effects across values of different attributes of the same (target) stimulus. For example, we show that, when individuals engage in interattribute comparisons, target values are perceived to be smaller as reference values associated with the same stimulus increase. This result can be described as a between-attributes and within-stimulus contrast effect. Furthermore, we also show that, when individuals make interattribute inferences, target values may be perceived to be larger as reference values associated with the same stimulus increase. This result may be seen as a between-attributes and within-stimulus assimilation (or anchoring) effect. Thus, our research provides evidence for assimilation and contrast effects between different attributes, where the direction of these effects is contingent on the evaluability of the target value and the compatibility of the target value with a value on another (reference) attribute.

Directions for Future Research

In this research, we provided evidence that individuals may rely on two distinct processes when evaluating a target value in the absence of same-attribute information. First, an evaluator may compare the target attribute value to the stimulus's value on a reference attribute (i.e., interattribute comparison process). Second, the evaluator may infer the magnitude of the target value based on her judgment about the extremity (e.g., the goodness or badness) of the stimulus's value on a reference attribute and the perceived correlation between the target attribute and the reference attribute (i.e., interattribute inference process). We argued and provided evidence that whether the evaluator relies on an interattribute comparison versus inference process is contingent on the evaluability of the target attribute and on the compatibility of the target with the reference value. We do believe that our research raises plenty of interesting questions for future research.

First, another potential moderator of our effects may be the perceived variability of the attributes. An attribute may vary from day to day for the same person, or may be relatively fixed within the same person. It seems possible that when performance is expected to be relatively stable over time and the correlation between attributes is high (e.g., between GPA and GMAT score), an interattribute inference process is more likely than when performance is highly variable (e.g., points and assists in a basketball game).[2]

Second, Study 5 presents evidence that the perceived correlation between the reference and the target moderates the strength of the interattribute inference process. It is certainly plausible that the perceived correlation between the reference and the target also moderates the strength of the interattribute comparison process. For example, it is plausible that to the extent that the evaluator believes that the height and the weight of an object are positively correlated (e.g., as in the case of TVs in Study 6), she perceives an object to be heavier as its height decreases while holding its actual weight constant. Conversely, if the reference attribute is seen as unrelated to the target attribute, the reference attribute might be seen as irrelevant as a point of comparison for the target attribute.

Finally, future research could also examine how individuals evaluate a target when multiple reference attributes are available. For instance, which factors determine how each reference attribute is weighted when evaluating the target? And to what extent does the degree to which the various reference attributes concur with each other influence the direction and magnitude of their effect on the target? Future research could expand our findings to situations where multiple reference (and possibly target) attributes are present. We believe that our research is only the first step in understanding individuals' interattribute evaluation processes.

Context of the Research

Our goal in this research project was to deepen our understanding of how we evaluate attribute values—a cornerstone of human

[2] We thank an anonymous reviewer for suggesting this potential moderator.

judgment and decision-making. The first author is particularly passionate about understanding and explaining humans' processing and use of attribute information. In the beginning of this research, he conducted some exploratory studies where he consistently found evidence of the interattribute inference process reported in this article. Later on, he presented those findings to the second author, who in turn suggested that humans may often engage in an interattribute comparison process that is directionally opposite to the inference process. That possibility was fascinating, and we embarked on a research journey together to try and explain when each of these processes occurs. We do acknowledge that our work may be—at least to some extent—incomplete, yet we do hope that it will inspire future research on what we feel is a fascinating topic.

References

Beckwith, N. E., & Lehmann, D. R. (1975). The importance of halo effects in multi-attribute models. *Journal of Marketing Research, 12,* 265–275. http://dx.doi.org/10.2307/3151224

Bless, H., & Schwarz, N. (2010). Mental construal and the emergence of assimilation and contrast effects: The inclusion/exclusion model. *Advances in Experimental Social Psychology, 42,* 319–373. http://dx.doi.org/10.1016/S0065-2601(10)42006-7

Fischer, G. W., & Hawkins, S. A. (1993). Strategy compatibility, scale compatibility, and the prominence effect. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 580–597. http://dx.doi.org/10.1037/0096-1523.19.3.580

Frederick, S. W., & Mochon, D. (2012). A scale distortion theory of anchoring. *Journal of Experimental Psychology: General, 141,* 124–133. http://dx.doi.org/10.1037/a0024006

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis*. New York, NY: Guilford Press.

Hsee, C. K. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes, 67,* 247–257. http://dx.doi.org/10.1006/obhd.1996.0077

Hsee, C. K., Loewenstein, G. F., Blount, S., & Bazerman, M. H. (1999). Preference reversals between joint and separate evaluation of options: A review and theoretical analysis. *Psychological Bulletin, 125,* 576–590. http://dx.doi.org/10.1037/0033-2909.125.5.576

Hsee, C. K., & Zhang, J. (2010). General evaluability theory. *Perspectives on Psychological Science, 5,* 343–355. http://dx.doi.org/10.1177/1745691610374586

Jaccard, J., & Wood, G. (1988). The effects of incomplete information on the formation of attitudes toward behavioral alternatives. *Journal of Personality and Social Psychology, 54,* 580–591. http://dx.doi.org/10.1037/0022-3514.54.4.580

Judd, C. M., Kenny, D. A., & McClelland, G. H. (2001). Estimating and testing mediation and moderation in within-subject designs. *Psychological Methods, 6,* 115–134. http://dx.doi.org/10.1037/1082-989X.6.2.115

Kardes, F. R., Posavac, S. S., & Cronley, M. L. (2004). Consumer inference: A review of processes, bases, and judgment contexts. *Journal of Consumer Psychology, 14,* 230–256. http://dx.doi.org/10.1207/s15327663jcp1403_6

Markman, A. B., & Loewenstein, J. (2010). Structural comparison and consumer choice. *Journal of Consumer Psychology, 20,* 126–137. http://dx.doi.org/10.1016/j.jcps.2010.01.002

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology, 35,* 250–256. http://dx.doi.org/10.1037/0022-3514.35.4.250

Sanbonmatsu, D. M., Vanous, S., Hook, C., Posavac, S. S., & Kardes, F. R. (2011). Whither the alternatives: Determinants and consequences of selective versus comparative judgemental processing. *Thinking & Reasoning, 17,* 367–386. http://dx.doi.org/10.1080/13546783.2011.625659

Shafir, E. (1995). Compatibility in cognition and decision. *Psychology of Learning and Motivation, 32,* 247–274. http://dx.doi.org/10.1016/S0079-7421(08)60312-6

Sherif, M., & Hovland, C. I. (1961). *Social judgment*. New Haven, CT: Yale University Press.

Sherif, M., Taub, D., & Hovland, C. I. (1958). Assimilation and contrast effects of anchoring stimuli on judgments. *Journal of Experimental Psychology, 55,* 150–155. http://dx.doi.org/10.1037/h0048784

Simonson, I., Bettman, J. R., Kramer, T., & Payne, J. W. (2013). Comparison selection: An approach to the study of consumer judgment and choice. *Journal of Consumer Psychology, 23,* 137–149. http://dx.doi.org/10.1016/j.jcps.2012.10.002

Slovic, P., Griffin, D., & Tversky, A. (1990). Compatibility effects in judgment and choice. In R. M. Hogarth (Ed.), *Insights in decision making* (pp. 5–27). Illinois: University of Chicago Press.

Slovic, P., & MacPhillamy, D. (1974). Dimensional commensurability and cue utilization in comparative judgment. *Organizational Behavior and Human Performance, 11,* 172–194. http://dx.doi.org/10.1016/0030-5073(74)90013-0

Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology, 4,* 25–29. http://dx.doi.org/10.1037/h0071663

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131. http://dx.doi.org/10.1126/science.185.4157.1124

Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review, 95,* 371–384. http://dx.doi.org/10.1037/0033-295X.95.3.371

Received April 6, 2018
Revision received November 21, 2018
Accepted November 21, 2018 ■